

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

Jc675 U.S. PTO
09/731773
12/08/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application:

1 9 9 9 年 1 2 月 9 日

出 願 番 号
Application Number:

平成 1 1 年 特 許 願 第 3 4 9 8 0 9 号

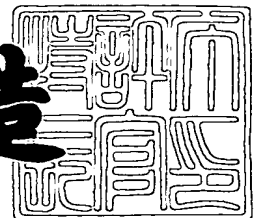
出 願 人
Applicant (s):

甲府日本電気株式会社

2 0 0 0 年 9 月 2 2 日

特 許 庁 長 官
Commissioner,
Patent Office

及 川 耕 造



出 証 番 号 出 証 特 2 0 0 0 - 3 0 7 8 1 5 1

【書類名】 特許願

【整理番号】 03905037

【提出日】 平成11年12月 9日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/00

【発明者】

 【住所又は居所】 山梨県甲府市大津町 1 0 8 8 - 3 甲府日本電気株式会
社内

 【氏名】 近藤 秀俊

【特許出願人】

 【識別番号】 000168285

 【氏名又は名称】 甲府日本電気株式会社

 【代表者】 桑田 幹雄

【代理人】

 【識別番号】 100105810

 【弁理士】

 【氏名又は名称】 根本 宏

【手数料の表示】

 【予納台帳番号】 072627

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9912421

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 ネットワークシステムにおけるデータアクセス方法、ネットワークシステムおよび記録媒体

【特許請求の範囲】

【請求項 1】 複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項 2】 複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりも読み出し速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一

のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項 3】 複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法。

【請求項 4】 複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備える共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリか

らデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステム。

【請求項 5】 複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備えると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一データがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステム。

【請求項 6】 請求項 4 および 5 のいずれかに記載のネットワークシステムにおいて、

前記タグメモリを前記通信機構に設けたことを特徴とするネットワークシステム

ム。

【請求項 7】 複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法を実現するためのデータアクセス用プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスする処理と、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する処理と、を含む処理をコンピュータに実行させるためのデータアクセス用プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ネットワークシステム、特に NUMA (Non Uniform Memory Access model) システムに適用して好適なデータアクセス技術に関する。

【0002】

【従来の技術】

図 21 は、NUMA システムの一例のブロック構成図であり、このシステムでは、複数のノード装置 1000a、…、1000n の夫々が、互いに所要の情報を通信可能にバス 1040 で接続されていて、各ノード装置は、CPU 1010a (…、1010n) と、ローカルメモリ 1020a (…、1020n) と、制御ユニット 1030a (…、1030n) とを有している。また、図 22 に示すように、各ノード装置のローカルメモリ 1020a、…、1020n は、共有メモリを構成していて、各ノード装置の CPU 1010a、…、1010n は、この共有メモリをアクセス可能に構成されている。

【0003】

そして、このようなNUMAシステムにおいては、各ノード装置 1 0 0 a、…、1 0 0 n 内のCPU 1 0 1 0 a、…、1 0 1 0 n から共有メモリへアクセス動作を行なう場合、読み出し開始からデータ読み出しが完了するまでの期間であるレイテンシが異なってしまうため、CPU 1 0 1 0 a、…、1 0 1 0 n の夫々が、キャッシュを内蔵することで、自ノード装置（以下、ローカルノードと称する場合もある）が、他ノード装置（以下、リモートノードと称する場合もある）のCPU内のキャッシュへアクセスする回数を減らすように工夫を施している。

【発明が解決しようとする課題】

しかしながら、共有メモリに記憶されているデータをキャッシュメモリにコピーする際に特に制約がない場合にあっては、共有メモリにおける同一アドレスのデータコピーが、複数のプロセッサ内のキャッシュメモリ内に保持されてしまい、ノード装置間におけるアクセスが頻繁に発生する。したがって、リモートノードのキャッシュメモリやローカルメモリに対するアクセス時のレイテンシ悪化がシステム性能低下をもたらしていた。

【0 0 0 4】

また、このシステムにあっては、共有メモリを構成する各ローカルメモリが、ノード装置間に分割して実装されるため、ローカルノードとリモートノードでのノード間のレイテンシ差が大きくなり、これもシステム性能低下をもたらしていた。

【0 0 0 5】

本発明は、このような従来の課題を鑑みてなされたもので、レイテンシを短縮できるデータアクセス方法、ネットワークシステム、データアクセス用プログラムを記録した記録媒体を提供することを目的としている。また、本発明の他の目的は、ノード間のレイテンシ差を小さくすることにある。

【課題を解決するための手段】

上記目的を達成するために、本発明の内の請求項 1 に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に

構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定することを特徴とした。

【0006】

この請求項 1 に係る発明においては、タグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、読み出したタグ情報に基づいて、投機アクセスによりメモリから獲得されたデータを廃棄するか否かを決定する。したがって、従来のように、まず、タグ情報を読み出してこの読み出したタグ情報が、システム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、さらに、メモリアクセス動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行してメモリアクセスを行なうにおいて、読み出したタグ情報がシステム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、既にアクセスしておいたデータを採用するので、レイテンシを短縮することができる。

【0007】

また、請求項 2 に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりも読み出し速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法である。

【0008】

この請求項2に係る発明においては、各ノード装置は、タグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送る。一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【0009】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、自ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行して自ノード装置内のメモリから投機データ読み出しを行なうにおいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出しをしておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【0010】

また、請求項3に係る発明は、複数のノード装置の夫々が互いに所要の情報を

通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法において、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しし、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する、ことを特徴とするネットワークシステムにおけるデータアクセス方法である。

【0011】

この請求項3に係る発明によれば、各ノード装置は、タグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【0012】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを

示す場合には、さらに、他ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行して他ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【0013】

また、請求項4に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備える共に、各ノード装置は、前記複数のノード装置内に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機読み出しする投機的読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステムである。

【0014】

この請求項4に係る発明においては、投機読み出し手段は、タグ情報をタグメモリから読み出すのと並行して、自ノード装置内のメモリからデータを投機的に読み出す。そして、読み出しデータ処理手段は、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても

存在しないことを示す場合には、この投機読み出しデータを自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【 0 0 1 5 】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、自ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行して自ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【 0 0 1 6 】

また、請求項 5 に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能とする通信機構を備えたと共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリからデータを読み出して、所定の処理を実行することが可能に構成されたネットワークシステムにおいて、

各ノード装置は、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする投機読み出し手段と、

読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、

一方、前記読み出したタグ情報が、前記投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、

このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、前記投機読み出しデータを廃棄する読み出しデータ処理手段と、を備えたことを特徴とするネットワークシステムである。

【0017】

この請求項5に係る発明においては、投機読み出し手段は、タグ情報をタグメモリから読み出すのと並行して、他ノード装置内のメモリからデータを投機読み出しする。そして、読み出しデータ処理手段は、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置内のプロセッサに送り、一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれかのキャッシュメモリにおいて存在することを示す場合には、このキャッシュメモリに存在するデータを獲得して自ノード装置内のプロセッサに送り、投機読み出しデータを廃棄する。

【0018】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データがいずれのキャッシュメモリにも存在しないことを示す場合には、さらに、他ノード装置内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行して他ノード装置内のメモリから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがいずれのキャッシュメモリにおいても存在しないことを示す場合には、既に投機読み出ししておいたデータをプロセッサに送るので、レイテンシを短縮することができる。

【0019】

また、請求項6に係る発明は、請求項4および5のいずれかに記載のネットワークシステムにおいて、

前記タグメモリを前記通信機構に設けたことを特徴とする。

【0020】

この請求項6に係る発明においては、上記効果に加えて、タグメモリを通信機構に設けることによって、いずれのノード装置からのタグ読み出しも略等しい時

間で行なえるので、ノード装置間のレイテンシ差を小さくすることができる。

【0021】

また、請求項7に係る発明は、複数のノード装置の夫々が互いに所要の情報を通信可能に構成されると共に、各ノード装置は、前記複数のノード装置に設けられたメモリやこのメモリよりもアクセス速度が高速なキャッシュメモリをアクセスして、所定の処理を実行することが可能に構成されたネットワークシステムにおけるデータアクセス方法を実現するためのデータアクセス用プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

システム内に設けられたキャッシュメモリのデータ格納状態に関する情報であるタグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスする処理と、

読み出したタグ情報に基づいて、前記投機アクセスにより前記メモリから獲得されたデータを廃棄するか否かを決定する処理と、を含む処理をコンピュータに実行させるためのデータアクセス用プログラムを記録した記録媒体である。

【0022】

この請求項7に係る発明においても、タグ情報をタグメモリから読み出すのと並行してシステム内のメモリを投機アクセスし、読み出したタグ情報に基づいて、投機アクセスによりメモリから獲得されたデータを廃棄するか否かを決定する。したがって、従来のように、まず、タグ情報を読み出してこの読み出したタグ情報が、システム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、さらに、メモリアクセス動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行してメモリアクセスを行なっていて、読み出したタグ情報がシステム内に設けられたキャッシュメモリにアクセス対象のデータが記憶されていないことを示す場合には、既にアクセスしておいたデータを採用するので、レイテンシを短縮することができる。

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照しつつ説明する。以下、本発明の理解の容易化のため2台のノード装置でNUMAシステムを構成する場合を例にとっ

て説明するが、3台以上のノード装置でNUMAシステムを構成するようにしても良い。図1に示すネットワークシステムは、2台のノード装置100a、100bを相互に所要の情報を通信可能にICN（Internal Connection Network：インターナルコネクションネットワーク）300で接続している。

【0023】

両ノード装置100a、100bは同一の構成であるので、ノード装置100a側のみの構成を説明すると、このノード装置100aは、メモリユニット（MU）30aと、このメモリユニットよりもアクセス速度が高速なキャッシュメモリ20aを内蔵する1台のプロセッサ10aと、アクセス制御等を行なうシステムコントロールユニット（SCU）200aとを備えていて、プロセッサ10aは、ライトバックやライトスルー等の公知のアルゴリズムによってキャッシュコヒーレンスを保証するように構成されている。

【0024】

システムコントロールユニット（SCU）200aは、プロセッサ10aとの間でインターフェイス制御動作を行うPIU（Processor Interface Unit：プロセッサインターフェイスユニット）40aと、MU（Memory Unit：メモリユニット）30aとの間でインターフェイス制御動作を行うMIU（Memory Interface Unit：メモリインターフェイスユニット）35aと、他ノード装置100bと間でインターフェイス制御動作を行うSIU（System Interface Unit：システムインターフェイスユニット）45aとを有していて、PIU40aは、プロセッサ10aから発行されるトランザクションのアドレスに基づいて、ホームノード（このアドレスが存在するノード装置）を検出する機能を有する。そして、図4に示すように、ノード装置100aのメモリユニット（MU）30aと、ノード装置100bのメモリユニット（MU）30bとで共有メモリを構成していて、両ノード装置100a、100bのプロセッサ10a、10bは、この共有メモリをアクセス可能に構成されている。

【0025】

通信機構としてのICN300は、データのルーティングを行うためのルーティング部320と、タグ情報を記憶するタグメモリ310とを備えている。図2

に示すように、タグ情報は、ブロックデータの番号（ブロック番号）と、ステータス情報と、ノード装置番号とが対となっている。図 3 に示すように、ステータス情報「U」は、いずれのノード装置 1 0 0 a、1 0 0 b 内のキャッシュメモリ 2 0 a、2 0 b においても、アクセス対象となるデータが無いことを示すもので、この場合ノード装置番号は与えられない。ステータス情報「S」は、キャッシュメモリ内のデータはメモリユニット内のデータと一致しており、かつ、複数のノード装置（この場合、例えば 1 0 0 a、1 0 0 b）のキャッシュメモリ 2 0 a、2 0 b においてアクセス対象となるデータと同一のデータが記憶されていることを示すもので、この場合には当該キャッシュデータを持つノード装置の番号がノード装置番号として対応付けられる。そして、ステータス情報「P」は、1 つのノード装置（この場合、1 0 0 a または 1 0 0 b）にのみ、キャッシュデータが有ることを示すもので、この場合には、このキャッシュデータを持つノード装置の番号がノード装置番号として対応付けられる。

【 0 0 2 6 】

図 5 は、本発明の主要部の構成図である。P I U 4 0 a は、プロセッサ 1 0 a に送るリプライデータを複数エントリ分格納する R D B (Reply Data Buffer : リプライデータバッファ) 4 2 0 と、R D B 4 2 0 のエントリ数分だけ設けられた制御ユニット (W 1 ~ W n) が設けられた R D C B (Reply Data Control Buffer : リプライデータコントロールバッファ) 4 1 0 と、R D C B 4 1 0 から、プロセッサ 1 0 a へのリプライが許可されたエントリーのうちから、いずれか 1 つのエントリを調停出力して R D B 4 2 0 に格納させるリプライ調停部 4 3 0 とを備えている。

【 0 0 2 7 】

図 6 に示すように、R D C B 4 1 0 を構成する一つの制御ユニット（例えば W 1）は、ビットメモリ 4 1 1 と、V ビット生成部 4 1 2 とを備えていて、エントリ単位に設けられる。図 7 に示すように、ビットメモリ 4 1 1 に格納されるビットは 6 種類ある。「T V ビット (Transaction V bit)」は、P I U 4 0 a がプロセッサ 1 0 a からのリード系トランザクションを受け付け、この受け付けたリード系トランザクションを M I U、S I U へ発行したことを示すビット、「L V

ビット (Local Memory V bit) 」は、ローカルノード内のメモリユニット (MU) であるローカルメモリからリプライデータが返却されたことを示すビット (Local Memory V bit) 、 「HVビット (Home Node V bit) 」は、他ノード装置がホームノードであった場合、ホームノードのメモリユニット (MU) からリプライデータが返却されたことを示すビット、 「CVビット (Cache V bit) 」は、リモートノード内のキャッシュメモリからリプライデータが返却されたことを示すビット、 「JVビット (Judgement V bit) 」は、SIUから返却されるスヌープ結果を示すビット、 「EVビット (Entry V bit) 」は、LV、TV、HV、CV、JVビットに基づいて、対応するエントリーのRDB 420に格納するリプライデータをプロセッサ10aへ送信可能であることを示すビットである。なお、タグメモリ310からタグ情報を読み出す動作を「スヌープ動作」と称し、この読み出し結果を「スヌープ結果」と称している。

【0028】

Vビット生成部412は、LV、TV、HV、CV、JVビットに基づいてEVビットを生成する。しかし、このEVビット生成がPIU40aへ報告されるタイミングは、各エントリー間で調整されていないため、複数エントリに対して複数のEVビットが立った場合、リプライ調停部430が調停出力した、RDB 508のエントリーに格納されたリプライデータが、プロセッサ10aに送られると共に、Vビット生成部412は、RDCB 509のEVビットをリセットしてエントリの無効化を行うように構成されている。一方、スヌープ処理と並行して、CPU10aから受け付けたリード系トランザクションに基づいて、ローカルメモリであるMU30aやリモートノード100b内のMU30bを投機アクセスする。図5に示す例では、ローカルメモリであるMU30aを投機読み出しするための専用バスである投機リードバス500によって、「PIU40a→MIU35a→MU30a」の経路でローカルメモリであるMU30aからのデータ読み出しを行う。

【0029】

SIU45aは、PIU40aから発行されたリード系トランザクションに回答して行なわれるタグメモリ310へのスヌープ処理のスヌープ結果待ちデータ

や読み出したスヌープ結果を保持する S C B (Snoop Control Buffer : スヌープコントロールバッファ) 4 5 0 と、S C B 4 5 0 からスヌープ結果を読み出し、各トランザクションのスヌープ結果を生成して、これを判定情報として R D C B 4 1 0 へ出力する判定部 4 6 0 とを備えている。図 8、図 9 に示すように、S C B 4 5 0 には、エントリー数分の 2 種類の情報を格納する。この 2 種類の情報は、P I U 4 0 a から発行されたリード系トランザクションに応答して実行すべきスヌープ処理が待ち状態であることを示す「Vビット」と、タグメモリ 3 1 0 からのスヌープ処理結果を表す「S N P 情報」である。Vビットが立った状態でスヌープ結果が得られると、判定部 4 6 0 によって、対象トランザクションの R D C B 4 1 0 内のエントリーに対応する判定情報を発行し、J V ビットの設定を行う。なお、判定部 4 6 0 で判定された S C B 4 5 0 内のエントリーの V ビット、S N P 情報をクリアして、S I U 4 5 a 内のトランザクション処理を完了するように構成されている。

【 0 0 3 0 】

また、判定部 4 6 0 は、タグメモリ 3 1 0 のスヌープ結果に基づいて判定情報を生成、出力するが、その判定情報は、キャッシュコヒーレンシのプロトコルに依存する。例えば、一般的な「M E S I」プロトコルを用いた場合、R D C B 5 0 9 に送る判定情報としては、リモートノード 1 0 0 b のキャッシュメモリ 2 0 b が Modify Line でデータを保持していることを示す「Modify 有」（キャッシュメモリにデータが有る）と、リモートノードのキャッシュが Modify Line でデータ保持していないことを示す「Modify 無」（キャッシュメモリにデータが無い）の 2 種類となる。

【 0 0 3 1 】

（第 1 の動作例）

次に動作説明をする。まず、図 1 0 のケース①、即ち、ノード装置 1 0 0 a をホームノードとし、ローカルメモリ (M U 3 0 a) に対するリード系トランザクションが、プロセッサ 1 0 a から発行された場合について、図 1 1 等を参照しつつ説明する。なお、この時、図 1 5 に示すようにタグメモリ 3 1 0 に記憶されるブロック番号 a のブロックデータに属するいずれかのデータをリード系トランザ

クシヨンの読み出し対象とする。

【0032】

図11の下側には、このケースの動作の処理手順を示すタイムチャートが示されている。図11に示すように、ノード装置100aのプロセッサ10aからリード系トランザクションが発行されると、PIU40aがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、ローカルメモリであるMU30aであることを把握する。PIU40aは、タグメモリ310に対するスヌープ処理をSIU45aに行なわせ、これと並行して投機リードパス500によって、MU30aへのリード制御をMIU35aに指示する（図11の符号A）。投機読み出しは、「PIU40a→MIU35a→MU30a→MIU35a→PIU40a」なる経路で、ローカルメモリであるMU35aからデータを読み出すことによって行なわれる。一方、PIU40aは、タグメモリ310に対するスヌープ処理を行うためにSIU45aに対してトランザクションを発行する。SIU45aは、タグメモリ310に対してスヌープ処理を行うためトランザクションをICN300経由で発行する。すると、ルーティング部320の動作によって、トランザクションは、「SIU45a→ICN300→タグメモリ310→ICN300→SIU45a」の経路で実行され、タグメモリ310からスヌープ情報を読み出す。SIU45aは、ステータス情報が「U」であるので、このトランザクションのスヌープ結果が「Modify無し」（即ち、いずれのキャッシュメモリにも存在しない）であると判断し、PIU40aに対してその旨の判定情報を送る。

【0033】

PIU40aでは、プロセッサ10aから発行されたリード系トランザクションをMIU35a、SIU45aに発行済みであること（TVビット）、MU30aからMIU35a経由でローカルメモリのデータが返却されていること（LVビット）、SIU15aからスヌープ結果としてトランザクションでアクセスするラインがリモートノード100bのキャッシュメモリ20bに「Modify」として登録されていないことを示す「Modify無し」の情報が報告されたこと、の3つ

の条件が満たされたことにより本トランザクションのリプライデータを、ローカルメモリ（MU30a）から読み出したデータであるとしてプロセッサ10aにリプライデータを送る。

【0034】

次に、上記リード系トランザクションにおける動作を、図5～図10等を参照しつつより詳細について説明する。CPU10aからリード系トランザクションが発行されると、PIU40aは、投機リードパス500を用いMIU35a経由でローカルメモリであるMU30aに対してリード系トランザクションを発行する。同時に、タグメモリ310に対するスヌープ処理を行うために、SIU45aに対してリード系トランザクションを発行する。PIU40aはMIU35a、SIU45aに対してトランザクションを発行したため、RDCB410の対応する制御ユニットのTVビットを立てる。MU30aに対して発行したトランザクションは、そのリプライデータをMIU35a経由でPIU40aに送る。PIU40aは、送られてきたリプライデータをRDB420の対応するエントリに登録すると共に、RDB420に登録したこのエントリに対するRDCB410の制御ユニットのLVビットを立てる。

【0035】

一方、SIU15aに対して発行したトランザクションは、タグメモリ310に対するスヌープ処理の結果が返却されるまで、SCB450内の対応するエントリのVビットを立てる。ローカルメモリの読み出しと並行して、タグメモリ310に対するスヌープ処理を行うため、「ICN300→タグメモリ310→ICN300」の経路でルーチング部がタグ情報を読み出し、読み出したタグ情報をSIU45aに送る。送られてきたタグ情報は、SCB450内の対応するエントリのSNBフィールドに登録される。そして、VビットとSNB情報とがそろると、判定部460はトランザクションのスヌープ結果を判断情報として出力し、RDCB410内の対応する制御ユニットのJVビットに、この判断情報を登録させる。この例では、タグ情報が「Modify無し」となっているため、リプライデータとしては、ローカルメモリ（MU30a）から読み出されたリードデータを採用すべきことが判断できる。PIU40aにおいては、リード系トランザ

クシヨンに対応したRDCB410のエントリ情報として、TVビット、LVビット、JVビットが全て立てられ、かつ、JVビットは「Modify無し」であることから、Vビット生成回路412はEVビットを立てる（図10の①参照）。EVビットが立ったエントリは、リプライ調停部430により調停出力され、これに対応するデータがRDB420からリプライデータとして出力され、リード系トランザクシヨン発行元のプロセッサ10aへリプライデータを送ることでリード系トランザクシヨンの処理が完了する。

【0036】

したがって従来の処理によれば、図11の下側図に示すように、まず、タグ情報を読み出し、これがステータス情報「U」であれば、今度は、ローカルメモリに対する読み出し動作を行っていたので、本発明の実施形態の投機読み出しを行なうことにより、レイテンシの短縮が図られることが分かる。

【0037】

（第2の動作例）

次に、図10のケース②、即ち、ノード装置100aをホームノードとし、ローカルメモリ（MU30a）に対するリード系トランザクシヨンが、プロセッサ10aから発行された場合について、図12等を参照しつつ説明する。なお、この時、図16に示すようにタグメモリ310に記憶されるブロック番号aのブロックデータに属するいずれかのデータをリード系トランザクシヨンの読み出し対象とし、ノード装置100bのノード装置番号を「②」とする。

【0038】

図12には、このケースの動作の処理手順を示すタイムチャートが示されている。図11に示すように、ノード装置100aのプロセッサ10aからリード系トランザクシヨンが発行されると、PIU40aがこれを受け付けつけて、トランザクシヨンのホームノードを検出するルーティング機能によって、トランザクシヨンでのアクセス対象となるメモリユニットが、ローカルメモリであるMU30aであることを把握する。PIU40aは、タグメモリ310に対するスヌープ処理をSIU45aに行なわせ、これと並行して投機リードパス500によって、MU30aへのリード制御をMIU35aに指示する（図11の符号B）。

投機読み出しは、「PIU40a→MIU35a→MU30a→MIU35a→PIU40a」なる経路でローカルメモリであるMU35aからデータを読み出すことによって行なわれる。一方、PIU40aは、タグメモリ310に対するスヌープ処理を行うためにSIU45aに対してトランザクションを発行する。SIU45aは、タグメモリ310に対してスヌープ処理を行うためトランザクションをICN300経由で発行する。すると、ルーティング部320の動作によって、トランザクションは、「SIU45a→ICN300→タグメモリ310→ICN300→SIU45a」の経路で実行され、タグメモリ310からスヌープ情報を読み出す。SIU45aは、ステータス情報が「P」であるので、このトランザクションのスヌープ結果が「Modify有り」（即ち、ノード装置100bのキャッシュメモリに存在する）であると判断し、他ノード装置100bのSIU45bに対して、キャッシュデータ読み出しのための制御情報を送る。そして、「SIU45b→PIU40b→プロセッサ10b→キャッシュメモリ20b→プロセッサ10b→PIU40b→SIU45b→ICN300」なる経路でキャッシュデータがICN300に送られ、ルーティング部320は、これをノード装置100aのSIU45aに送る。すると、SIU45aは、これをPIU40aに送る。

【0039】

PIU40aでは、プロセッサ10aから発行されたリード系トランザクションをMIU35a、SIU45aに発行済みであること（TVビット）、MU30aからMIU35a経由でローカルメモリのデータが返却されていること（LVビット）、リモートノード100bのキャッシュメモリ20bに記憶するキャッシュデータが送られてきたこと（CVビット）、SIU15aからスヌープ結果としてトランザクションでアクセスするラインがリモートノード100bのキャッシュメモリ20bに「Modify」として登録されていることを示す「Modify有り」の情報が報告されたこと、の4つの条件が満たされたことにより本トランザクションのリプライデータを、ノード装置100bのキャッシュメモリ20bから読み出したデータであるとしてプロセッサ10aにリプライデータを送る。なお、先にMU30aから読み出されたデータは廃棄される。

【0040】

このように第1および第2の動作例によれば、ノード装置100aは、タグ情報をタグメモリ310から読み出すのと並行して、自ノード装置100a内のメモリユニット(MU)30aからデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータ(例えば同一アドレスまたは同一アドレス範囲のデータ)が全キャッシュメモリ20a、20bにおいて存在しないことを示す場合には、この投機読み出しデータを自ノード装置100a内のプロセッサ10aに送る。一方、読み出したタグ情報が、投機読み出し対象データと同一のデータがキャッシュメモリ20aまたは20bにおいて存在することを示す場合には、このキャッシュメモリ20aまたは20bに存在するデータを獲得して自ノード装置100a内のプロセッサ10aに送り、投機読み出しデータを廃棄する。

【0041】

したがって、従来のように(図11上側図参照)、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データと同一のデータが全キャッシュメモリ10a、10bに存在しないことを示す場合には、さらに、自ノード装置100a内のメモリユニット(MU)30aからのデータ読み出し動作を行なうといった一連の動作を行なうことはせず、タグ情報の読み出しと並行して自ノード装置100a内のメモリユニット(MU)30aから投機データ読み出しを行なっておいて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータが全キャッシュメモリ20a、20bにおいて存在しないことを示す(例えばステータス情報「U」)場合には、既に投機読み出ししておいたデータをプロセッサ10aに送るので、レイテンシを短縮することができる。

【0042】

(第3の動作例)

次に、図10のケース③、即ち、他ノード装置100bをホームノードとし、メモリユニット(MU30b)に対するリード系トランザクションが、プロセッサ10aから発行された場合について、図13等を参照しつつ説明する。なお、この時、図17に示すようにタグメモリ310に記憶されるブロック番号bのブ

ロックデータに属するいずれかのデータをリード系トランザクションの読み出し対象とする。

【0043】

図13には、このケースの動作の処理手順を示すタイムチャートが示されている。図13に示すように、ノード装置100aのプロセッサ10aからリード系トランザクションが発行されると、PIU40aがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、リモートノード100b内のMU30bであることを把握する。PIU40aは、タグメモリ310に対するスヌープ処理をSIU45aに行なわせ、これと並行してMU30bに対する投機読み出しを行なうために、ICN300に対して制御情報を送る。ルーティング部320は、これをSIU45bに送る(図13に符号C)。投機読み出しは、「SIU45b→MIU35b→MU30b→MIU35b→SIU45b」なる経路でリモートノード100bのMU30bからデータを読み出すことによっ
て行なわれる。そして、この読み出しデータはICN300に送られ、そのルーティング部320の動きによってノード装置100aのSIU45aを経由して、PIU40aに送られる。

【0044】

一方、PIU40aは、タグメモリ310に対するスヌープ処理を行うためにSIU45aに対してトランザクションを発行する。SIU45aは、タグメモリ310に対してスヌープ処理を行うためトランザクションをICN300経由で発行する。するとルーティング部320の動作によって、トランザクションは、「SIU45a→ICN300→タグメモリ310→ICN300→SIU45a」の経路で実行され、タグメモリ310からスヌープ情報を読み出す。SIU45aは、ステータス情報が「U」であるので、このトランザクションのスヌープ結果が「Modify無し」(即ち、いずれのキャッシュメモリにも存在しない)であると判断し、PIU40aに対してその旨の判定情報を送る。

【0045】

PIU40aでは、プロセッサ10aから発行されたリード系トランザクシ

ンをMIU35a、SIU45aに発行済みであること(TVビット)、他ノード装置のMU30bからデータが返却されていること(HVビット)、SIU15aからスヌープ結果としてトランザクションでアクセスするラインがリモートノード100bのキャッシュメモリ20bに「Modify」として登録されていないことを示す「Modify無し」の情報が報告されたこと、の3つの条件が満たされたことにより本トランザクションのリプライデータを、ローカルノード100bのMU30bから読み出したデータであるとしてプロセッサ10aにリプライデータを送る。

【0046】

(第4の動作例)

次に、図10のケース④、即ち、他ノード装置100bをホームノードとし、メモリユニット(MU30b)に対するリード系トランザクションが、プロセッサ10aから発行された場合について、図14等を参照しつつ説明する。なお、この時、図18に示すようにタグメモリ310に記憶されるブロック番号bのブロックデータに属するいずれかのデータをリード系トランザクションの読み出し対象とし、ノード装置100bのノード装置番号を「②」とする。

【0047】

図14には、このケースの動作の処理手順を示すタイムチャートが示されている。図14に示すように、ノード装置100aのプロセッサ10aからリード系トランザクションが発行されると、PIU40aがこれを受け付けつけて、トランザクションのホームノードを検出するルーティング機能によって、トランザクションでのアクセス対象となるメモリユニットが、ローカルノード100bのMU30bであることを把握する。PIU40aは、タグメモリ310に対するスヌープ処理をSIU45aに行なわせ、これと並行してMU30bに対する投機読み出しを行なうために、ICN300に対して制御情報を送る。ルーティング部320はこれをSIU45bに送る(図13に符号D)。投機読み出しは、「SIU45b→MIU35b→MU30b→MIU35b→SIU45b」なる経路でリモートノード100bのMU30bからデータを読み出すことによって行なわれる。そして、この読み出しデータはICN300に送られ、そのルーティ

ング部 3 2 0 の動作によってノード装置 1 0 0 a の S I U 4 5 a を経由して、P I U 4 0 a に送られる。

【0 0 4 8】

一方、S I U 4 5 a は、ステータス情報が「P」であるので、このトランザクションのスヌープ結果が「Modify有り」（即ち、ノード装置 1 0 0 b のキャッシュメモリに存在する）であると判断し、他ノード装置 1 0 0 b の S I U 4 5 b に対して、キャッシュデータ読み出しのための制御情報を送る。そして、「S I U 4 5 b → P I U 4 0 b → プロセッサ 1 0 b → キャッシュメモリ 2 0 b → プロセッサ 1 0 b → P I U 4 0 b → S I U 4 5 b → I C N 3 0 0」なる経路でキャッシュデータが I C N 3 0 0 に送られ、ルーティング部 3 2 0 は、これをノード装置 1 0 0 a の S I U 4 5 a に送る。すると、S I U 4 5 a は、これを P I U 4 0 a に送る。

【0 0 4 9】

P I U 4 0 a では、プロセッサ 1 0 a から発行されたリード系トランザクションを M I U 3 5 a、S I U 4 5 a に発行済みであること（TVビット）、他ノード装置の M U 3 0 b からデータが返却されていること（HVビット）、リモートノード 1 0 0 b のキャッシュメモリ 2 0 b に記憶するキャッシュデータが送られてきたこと（CVビット）、S I U 1 5 a からスヌープ結果としてトランザクションでアクセスするラインがリモートノード 1 0 0 b のキャッシュメモリ 2 0 b に「Modify」として登録されていることを示す「Modify有り」の情報が報告されたこと、の4つの条件が満たされたことにより本トランザクションのリプライデータを、ノード装置 1 0 0 b のキャッシュメモリ 2 0 b から読み出したデータであるとしてプロセッサ 1 0 a にリプライデータを送る。なお、先に M U 3 0 b から読み出されたデータは廃棄される。

【0 0 5 0】

したがって、第3および第4の動作例によれば、ノード装置 1 0 0 a は、タグ情報をタグメモリ 3 1 0 から読み出すのと並行して、他ノード装置 1 0 0 b 内の M U 3 0 b からデータを投機読み出しする。そして、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータ（例えば同一アドレスまたは同一ア

ドレス範囲のデータ) が全キャッシュメモリ 2 0 a、2 0 bにおいても存在しないことを示す場合には、この投機読み出しデータを獲得して自ノード装置 1 0 0 a 内のプロセッサ 1 0 a に送る。一方、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータがキャッシュメモリ 2 0 a または 2 0 b に存在することを示す場合には、このキャッシュメモリ 2 0 a または 2 0 b に存在するデータを獲得して自ノード装置 1 0 0 a 内のプロセッサ 1 0 a に送り、投機読み出しデータを廃棄する。

【0 0 5 1】

したがって、従来のように、まず、タグ情報を読み出し、この読み出したタグ情報が、読み出し対象データと同一のデータが全キャッシュメモリ 2 0 a、2 0 b に存在しないことを示す場合には、さらに、他ノード装置 1 0 0 b 内のメモリからのデータ読み出し動作を行なうといった一連の動作を行なうのではなく、タグ情報の読み出しと並行して他ノード装置 1 0 0 b 内の MU 3 0 b から投機データ読み出しを行なっていて、読み出したタグ情報が、投機読み出し対象となるデータと同一のデータが全キャッシュメモリ 2 0 a、2 0 b に存在しないことを示す(例えばステータス情報「U」) 場合には、既に投機読み出しておいたデータをプロセッサ 1 0 a に送るので、レイテンシを短縮することができる。

【0 0 5 2】

また、以上説明してきた実施の形態においては、タグメモリ 3 1 0 を通信機構としての ICN 3 0 0 に設けているので、いずれのノード装置 1 0 0 a、1 0 0 b からのタグ読み出しも略等しい時間で行なえるので、ノード装置間のレイテンシ差を小さくすることができる。

【0 0 5 3】

図 1 9 は、本発明の他の実施形態のネットワークシステムのブロック構成図である。この実施形態においては、タグメモリ 3 1 0 を ICN 3 0 0 に設けるのではなく、各ノード装置 1 0 0 a、1 0 0 b 内に分割して設けた点に特徴があり、その他の点は図 1 に示した構成例と変わる所がない。図 1 9 に示すように、ノード装置 1 0 0 a、1 0 0 b には、夫々、タグメモリ 3 1 1、3 1 2 が設けられており、この両タグメモリ 3 1 1、3 1 2 のタグ情報を纏めたものが、上述してき

たタグメモリ 310 のタグ情報と一致する。

【0054】

上述したケース①を例にとって、簡単に動作説明を行なう。プロセッサ 10a からリード系トランザクションが発行されると、PIU40a は SIU45a に対してスヌープ処理を実行させると共に、これと並行して MIU35a に対して投機読み出しを実行させる。投機読み出しは、「MIU35a→MU30a→MIU35a→PIU40a」の経路で行なわれて PIU は、ローカルメモリ (MU30a) のデータを得てスヌープ待ち状態となる。この時点で、RDCB410 の制御ユニットの LV ビット、TV ビットが立つことになる。

【0055】

一方、SIU45a は、2 つのタグメモリ 311、312 をスヌープする必要がある。したがって、タグメモリ 311 のステータス情報を獲得すると共に、ICN300 に対してスヌープ処理実行のための情報を送る。なお、ここまでの処理中に、図 20 (a) に示すように SCB450 の V ビットが立つ。次に、スヌープ処理実行のための情報は、ルーティング部 320 によって、ノード装置 100b の SIU45b に送られ、SIU45b は、タグメモリ 312 のスヌープ処理結果を ICN320 に送る。ルーティング部 320 は、これをノード装置 100a の SIU45a に送る。したがって、図 20 (b) に示すように V ビットが立つと共に、タグメモリ 311 のスヌープ結果を示す情報である SNP1 とタグメモリ 312 のスヌープ結果を示す情報である SNP2 とが登録されるので、判定部 460 は、これらの情報に基づいた判定結果を RDCB410 に送る。したがって、LV ビット、TV ビット、JV ビットがビットメモリ 411 に登録されるので V ビット生成部 412 は、EV ビットを立てる。リプライ調停部 430 は、EV ビットが立ったので、これに対応する RDB420 内のエントリをプロセッサ 10a に送ることになる。ステータス情報が「U」で、投機読み出し対象データと同一のデータがキャッシュメモリ 20a、20b に存在しない場合を想定すると、先に投機読み出しされたデータがリプライデータとしてプロセッサ 10a に送られることになる。

【0056】

このように、タグメモリ 310 を集中管理せずに、分割してネットワークシステムに分散するようにしても、投機読み出しによるレイテンシの短縮という同様の効果が得られることになる。また、タグメモリ 310 を通信機構内に設けなくても良いので通信機構の構成が簡素化される。

【0057】

なお、今まで説明した動作を実行するためのデータアクセス用プログラムをコンピュータ読み取り可能な記録媒体に記録しておいて、プロセッサ 10a (10b) がこれを読み取り実行することによって以上説明した動作を行なわせるようにすれば良い。このような記録媒体としては ROM 等の半導体記録媒体、FD、HD (ハードディスク) 等の磁気記録媒体、CDROM、DVDROM 等の光記録媒体等が挙げられる。もちろん、このようなアクセス用プログラムを、例えばハードディスク等で実現したローカルメモリ (MU30a, 30b) の所定領域に記録しておいて、プロセッサ 10a (10b) がこれを実行するようにしても良い。

【0058】

以上本発明の実施の形態について説明してきたが、本発明の要旨を逸脱しない範囲内で上述した実施形態に対して種々の変更や変形を施すことは可能であり、例えば、ノード装置の台数を 3 台以上にすること、各ノード装置内におけるプロセッサの台数を 2 台以上にすること、キャッシュメモリをプロセッサの外部に設けること等の変形が考えられることは言うまでもない。

【発明の効果】

以上説明してきたように、本発明によれば、投機読み出しをすることによって、ノード装置のキャッシュメモリやローカルメモリに対するアクセス時のレイテンシを短縮することが可能になるという効果が得られる。

【0059】

また、タグ情報を記憶するタグメモリを通信機構に備えるようにして、ノード間のレイテンシ差を小さくすることができるとい効果も得られる。

【図面の簡単な説明】

【図 1】

本発明の実施の形態であるネットワークシステムのブロック構成図である。

【図 2】

タグメモリ 3 1 0 に記憶されるタグ情報の説明図である。

【図 3】

ステータス情報の説明図である。

【図 4】

共有メモリのメモリマッピング状態を示す説明図である。

【図 5】

本発明の実施形態の主要部の構成図である。

【図 6】

R D C B 4 1 0 内の制御ユニットの構成図である。

【図 7】

ビットメモリ 4 1 1 に格納される各ビットの説明図である。

【図 8】

S C B 4 5 0 の構成図である。

【図 9】

S C B 4 5 0 において用いる制御情報の説明図である。

【図 1 0】

動作説明のための説明図である。

【図 1 1】

第 1 の動作例と従来技術とのレイテンシの差を説明するためのタイミングチャートである。

【図 1 2】

第 2 の動作例を説明するためのタイミングチャートである。

【図 1 3】

第 3 の動作例を説明するためのタイミングチャートである。

【図 1 4】

第 4 の動作例を説明するためのタイミングチャートである。

【図 1 5】

第 1 の動作例を説明するためのタグ情報の具体例の説明図である。

【図 1 6】

第 2 の動作例を説明するためのタグ情報の具体例の説明図である。

【図 1 7】

第 3 の動作例を説明するためのタグ情報の具体例の説明図である。

【図 1 8】

第 4 の動作例を説明するためのタグ情報の具体例の説明図である。

【図 1 9】

本発明の他の実施の形態のネットワークシステムのブロック構成図である。

【図 2 0】

他の実施の形態のネットワークシステムの動作を説明するための説明図である。

【図 2 1】

従来技術の説明図である。

【図 2 2】

従来技術の説明図である。

【符号の説明】

1 0 0 a、1 0 0 b ノード装置

1 0 a、1 0 b プロセッサ

2 0 a、2 0 b キャッシュメモリ

3 0 a、3 0 b MU (メモリユニット)

3 5 a、3 5 b MIU (メモリインターフェイスユニット)

4 0 a、4 0 b PIU (プロセッサインターフェイスユニット)

4 5 a、4 5 b SIU (システムインターフェイスユニット)

2 0 0 a、2 0 0 b SCU (システムコントロールユニット)

3 0 0 ICN (インターナルコネクションネットワーク)

3 1 0 タグメモリ

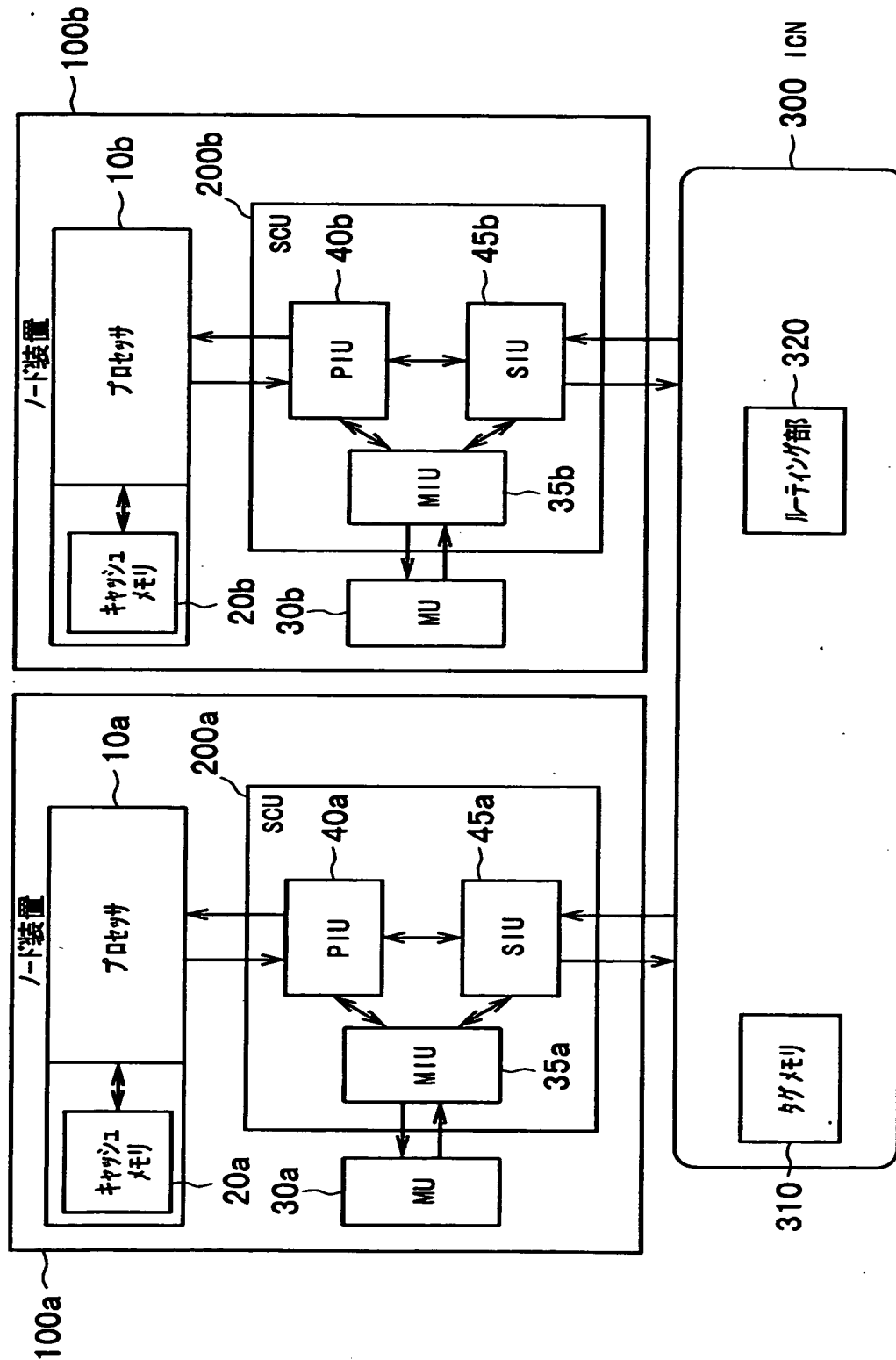
3 1 1 タグメモリ

3 1 2 タグメモリ

- 3 2 0 ルーティング部
- 4 1 0 RDCB (リプライデータコントロールバッファ)
- 4 1 1 ビットメモリ
- 4 1 2 Vビット生成部
- 4 2 0 RDB (リプライデータバッファ)
- 4 3 0 リプライ調停部
- 4 5 0 SCB (スヌープコントロールバッファ)
- 4 6 0 判定部

【書類名】 図面

【図 1】



【図 2】

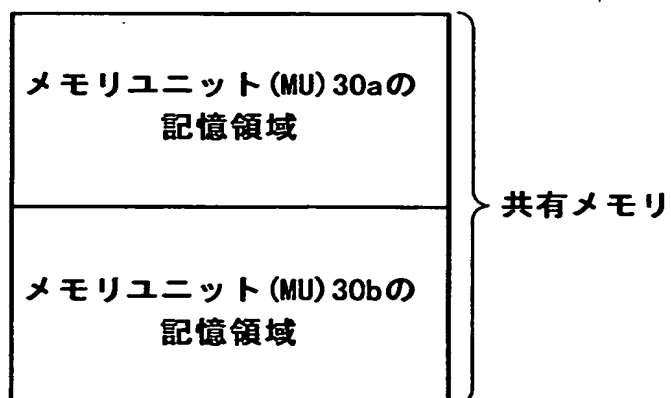
タグ情報

ブロック番号	ステータス情報	ノード装置番号
a	U	——
b	S	1, 2
⋮	⋮	⋮
n	P	2

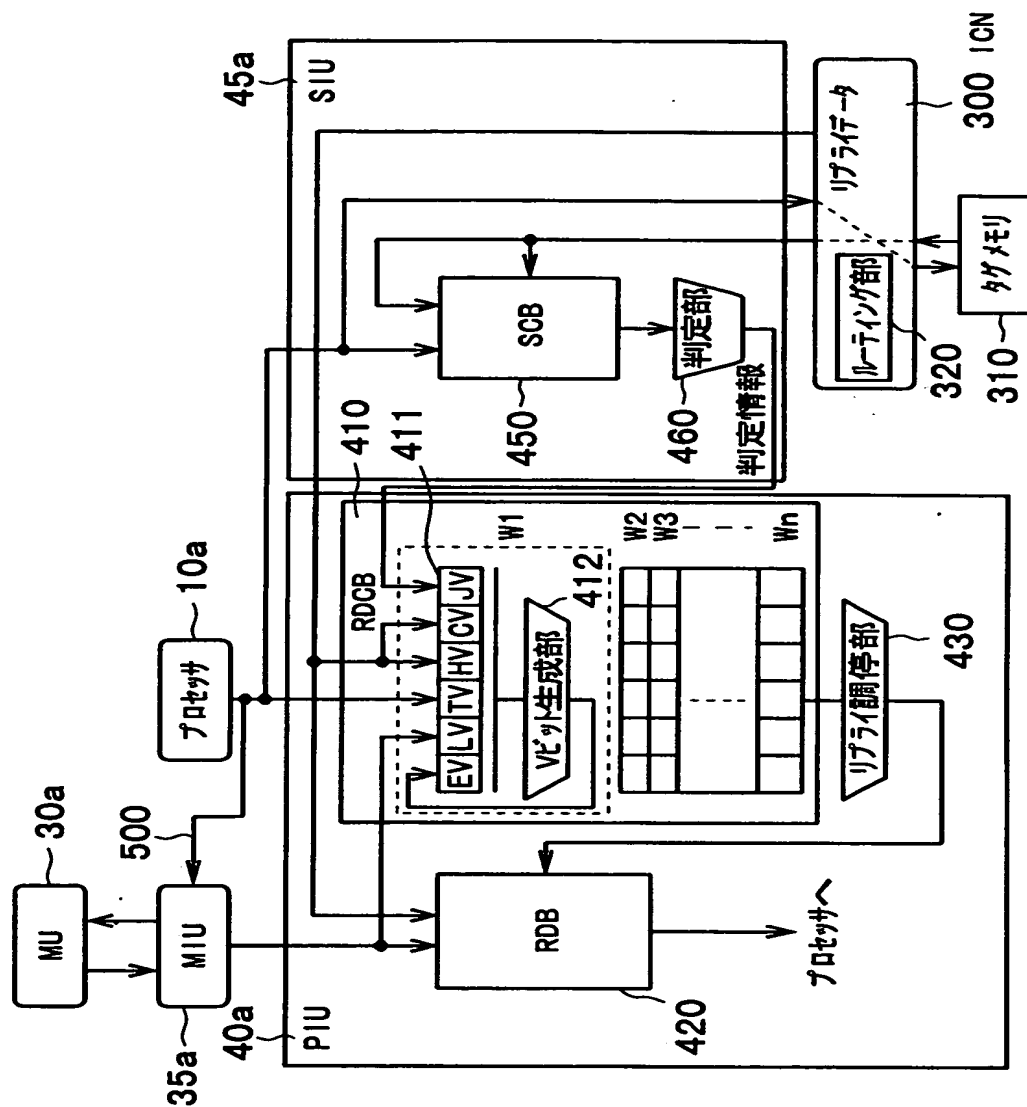
【図 3】

ステータス情報	キャッシュ状態
U	いずれのノード装置においても、キャッシュデータが無い
S	キャッシュ内のデータは、メモリユニット内のデータと一致し、かつ、複数のノード装置において同一キャッシュデータ有り
P	1つのノード装置にのみ、キャッシュデータが有る

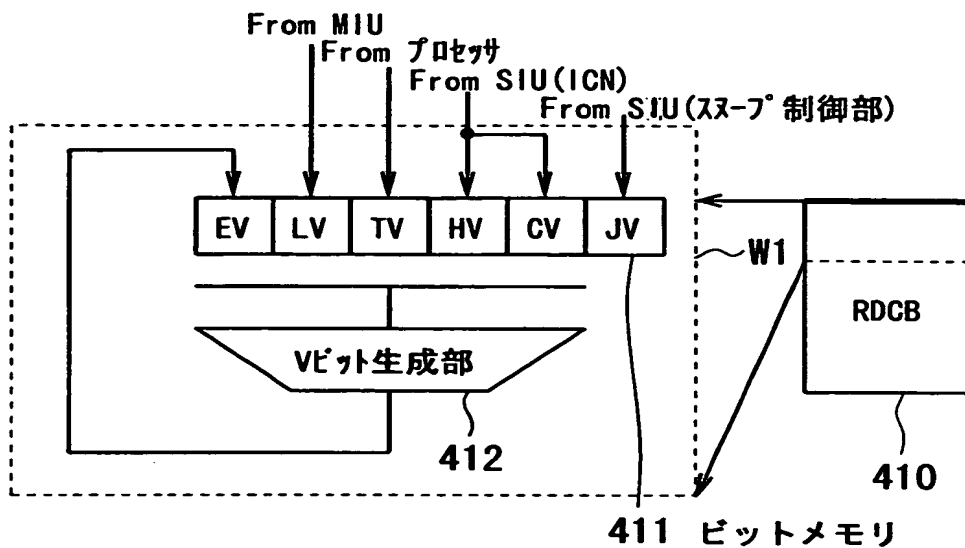
【図 4】



【図 5】



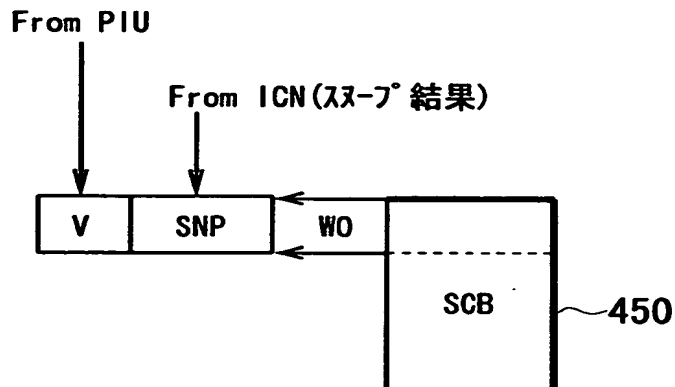
【図 6】



【図 7】

EV	プロセッサにリプライデータを送る条件が成立したことを示すエントリ代表ビット
LV	ローカルメモリから投機リードパスによるリプライデータが返却されたことを示すビット
TV	プロセッサからリード系トランザクションを受け付け、SIU/MIUに発行したことを示すビット
HV	他ノードに実装するメモリユニット(MU)からリプライデータが返却されたことを示すビット
CV	リモートノードのキャッシュに持つModifyデータが返却されたことを示すビット
JV	スヌープ結果を示すビット

【図 8】



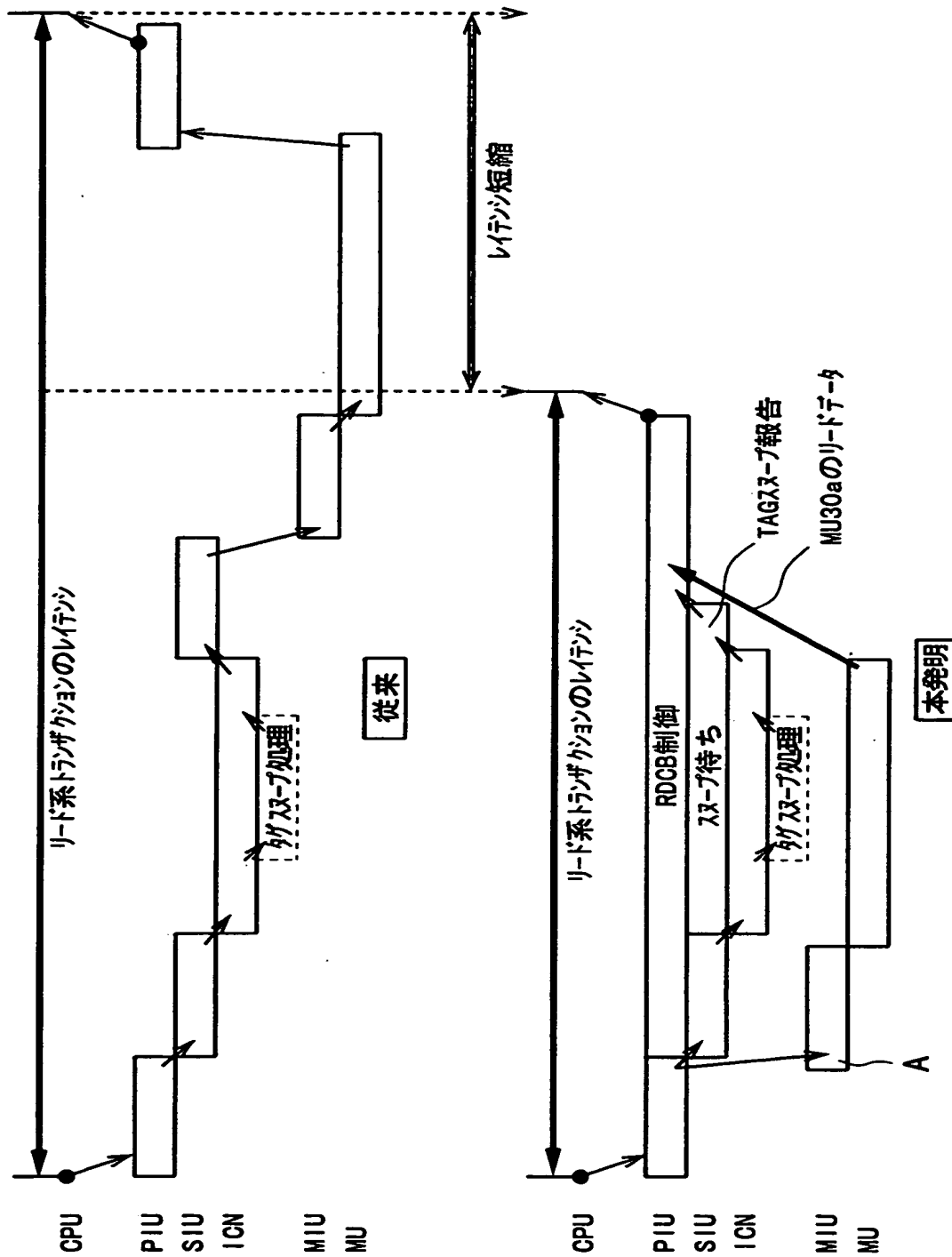
【図 9】

V	プロセッサからリード系トランザクションが発行され、スヌープ処理待ちであることを示すビット
SNP	Internal Connection Networkからのスヌープ結果を示す情報

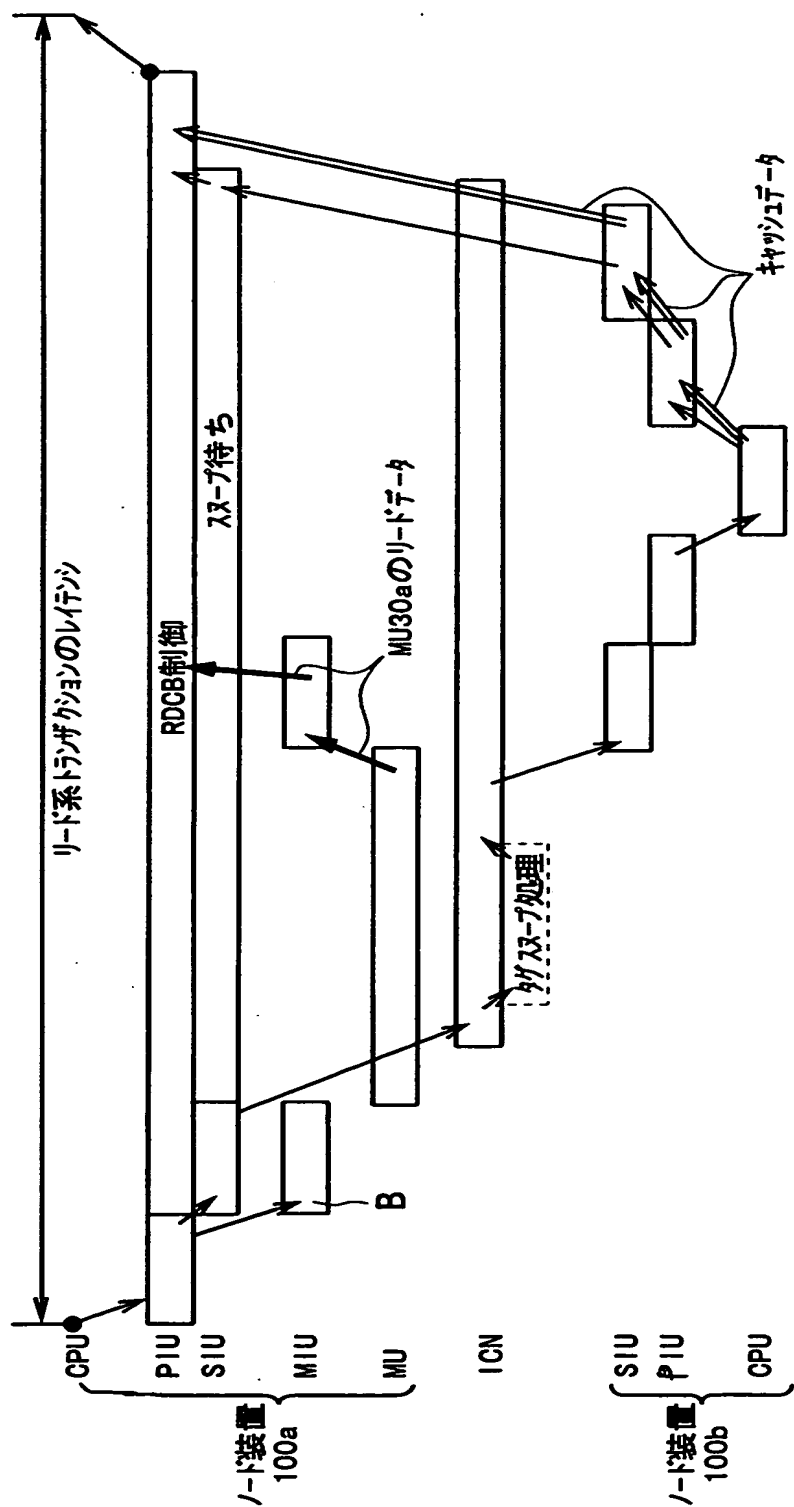
【図 1 0】

ケース	LV	TV	HV	CV	JV	処理内容
①	1	1	0	0	Modify 無	ローカルメモリデータをプロセッサに送る
②	1	1	0	1	Modify 有	キャッシュデータをプロセッサに送り、ローカルメモリデータ(投機読み出しデータ)を廃棄する
③	0	1	1	0	Modify 無	リモートノードのメモリデータをプロセッサに送る
④	0	1	1	1	Modify 有	キャッシュデータをプロセッサに送り、リモートノードのメモリデータを廃棄する

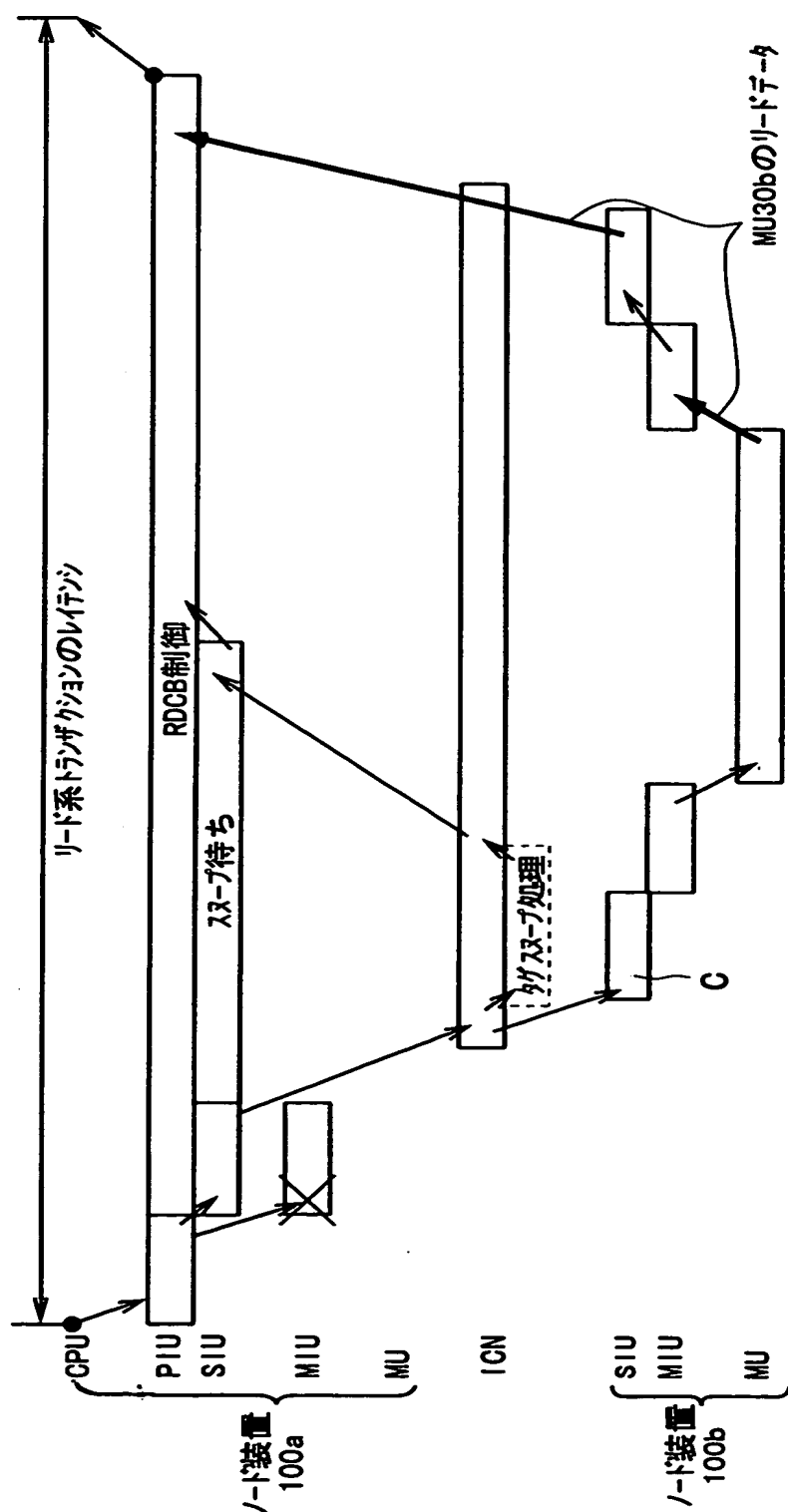
【図 11】



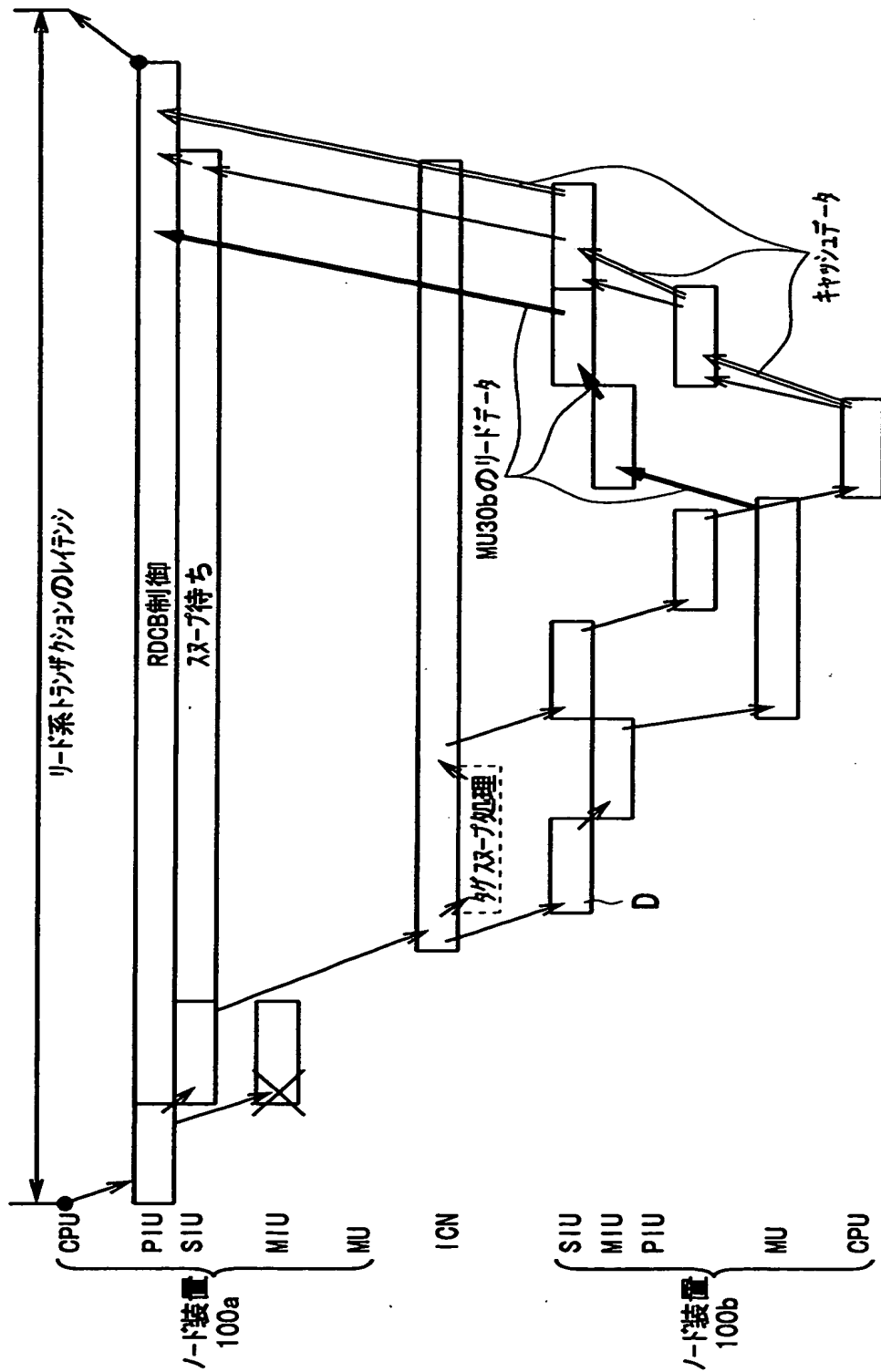
【图 1 2】



【图 13】



【图 14】



【図 1 5】

ブロック番号	ステータス情報	ノード装置番号
a	U	——
⋮	⋮	⋮

【図 1 6】

ブロック番号	ステータス情報	ノード装置番号
a	P	②
⋮	⋮	⋮

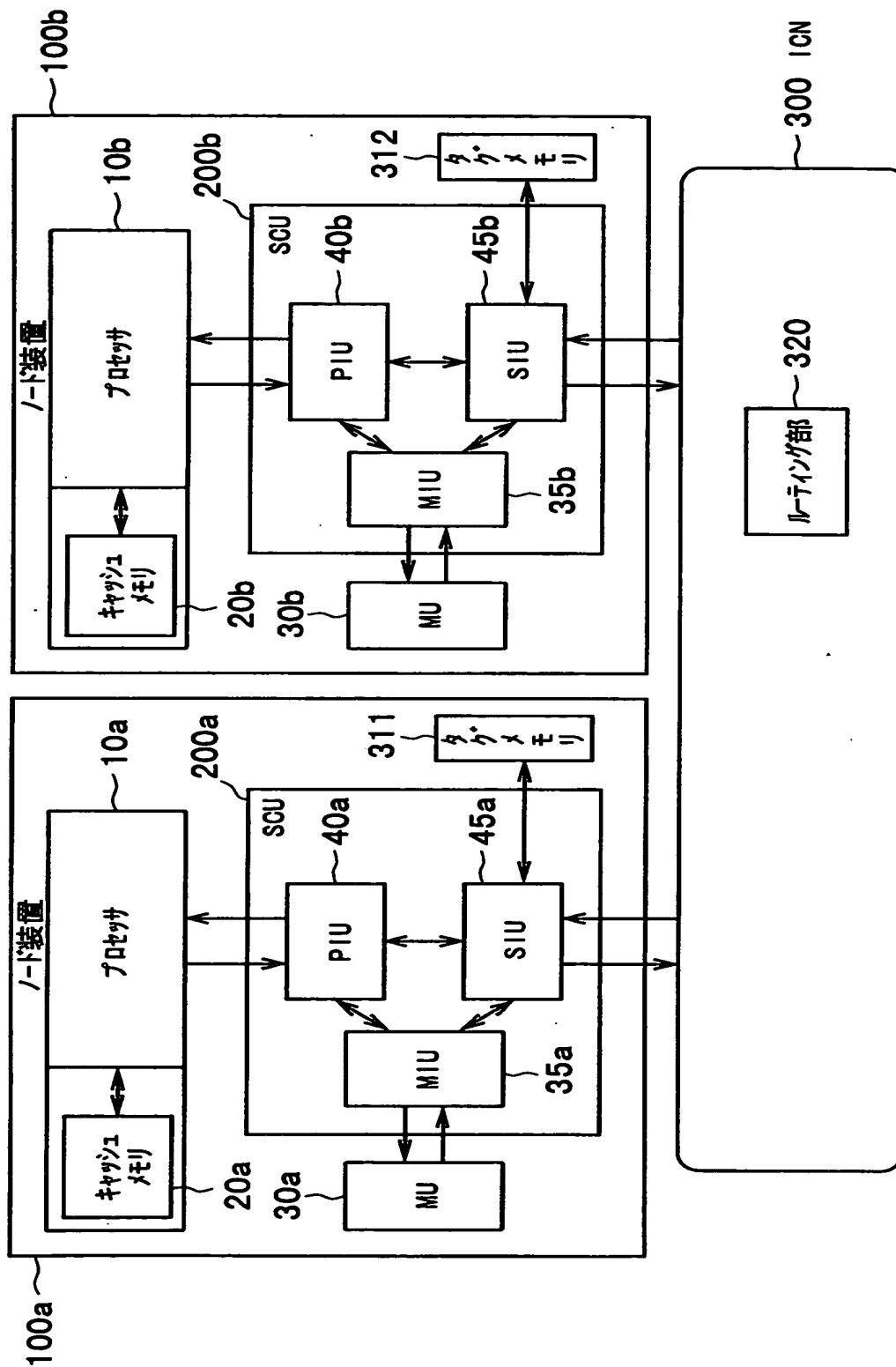
【図 1 7】

ブロック番号	ステータス情報	ノード装置番号
b	U	——
⋮	⋮	⋮

【図 1 8】

ブロック番号	ステータス情報	ノード装置番号
b	P	②
⋮	⋮	⋮

【図 1 9】



【図 2 0】

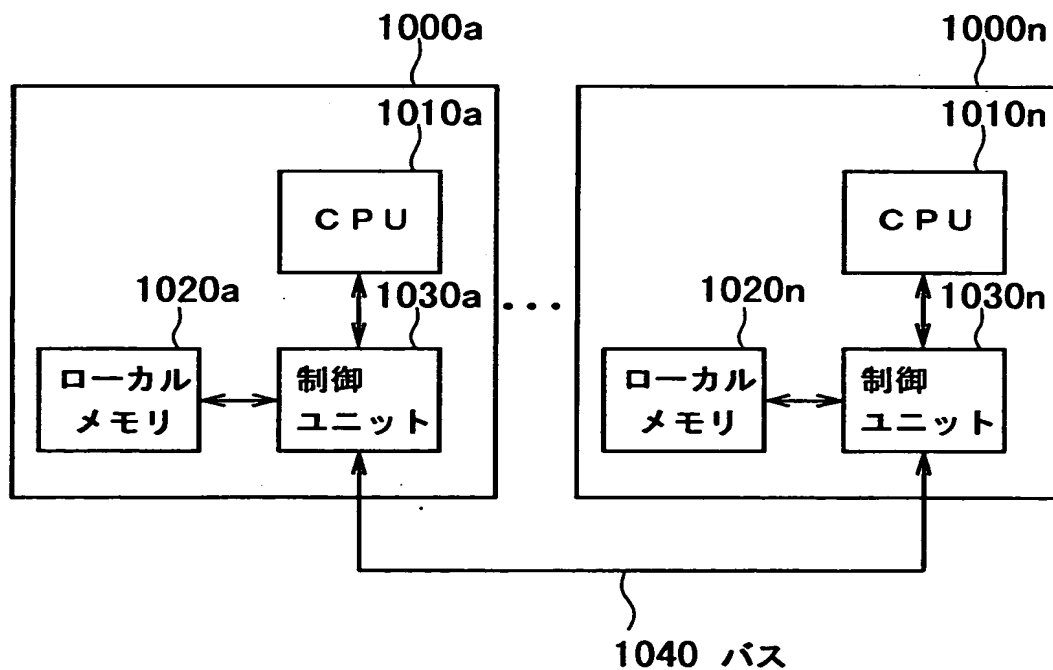
(a)

V	SNP 1	SNP 2
1		

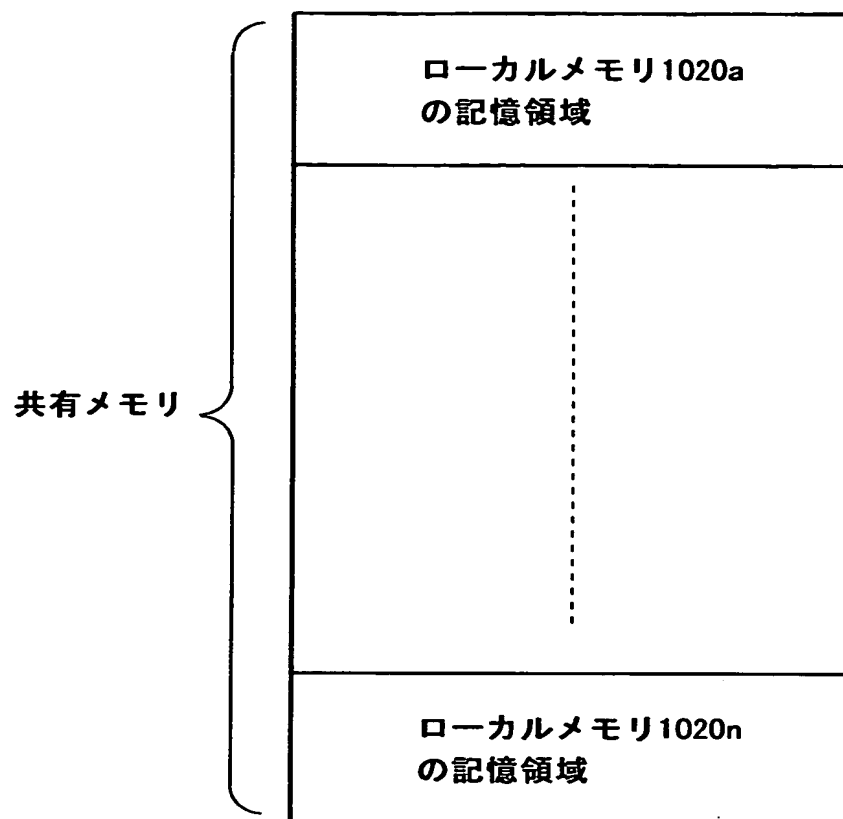
(b)

V	SNP 1	SNP 2
1	スヌープ結果	スヌープ結果

【図 2 1】



【図 2 2】



【書類名】 要約書

【要約】

【課題】 レイテンシを短縮できるデータアクセス方法を提供する。

【解決手段】 タグ情報をタグメモリ 3 1 0 から読み出すのと並行して、自ノード装置 1 0 0 a 内のメモリユニット (MU) 3 0 a からデータを投機読み出しする。そして、読み出したタグ情報が、投機的読み出しデータが全キャッシュメモリ 2 0 a、2 0 b において存在しないことを示す場合には、この投機読み出しデータを自ノード装置 1 0 0 a 内のプロセッサ 1 0 a に送る。一方、読み出したタグ情報が、投機読み出しデータがキャッシュメモリ 2 0 a または 2 0 b において存在することを示す場合には、このキャッシュメモリ 2 0 a または 2 0 b に存在するデータを獲得して自ノード装置 1 0 0 a 内のプロセッサ 1 0 a に送り、投機読み出しデータを廃棄する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000168285]

1. 変更年月日	1990年 8月 9日
[変更理由]	新規登録
住 所	山梨県甲府市大津町1088-3
氏 名	甲府日本電気株式会社